BGGN 213 Final Exam

Pierce Ford (PID: A59010464)

7/18/2022

Examining COVID Variant Prevalence Over Time

```
#Load in the necessary libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
#Load in the data
raw_data <- read.csv("covid19_variants.csv")
```

Prepping the Data

I'll remove the data for total and other across all variants, as I am only interested in the percentages for each named variant. I'll also convert the date column to actually be date objects for formatting later.

```r
#Filter out the total and other data points
filtered_data <- raw_data %>% filter((variant_name != "Total"
                                      & variant_name != "Other"))

#Convert first column to date objects for formatting x-axis labels later
filtered_data$date <- ymd(filtered_data$date)
```

Plotting the Data

I'll display a line chart comparing prevalence of each of the named COVID variants.

```r
#Draw chart with date on the x-axis and percentage on the y-axis, colored by
#variant name
ggplot(data = filtered_data, aes(x = date, y = percentage,
                                 color = variant_name)) +
  #Draw a line for each variant
  geom_line(aes(group = variant_name)) +
  #Change theme to black and white
  theme_bw() +
  #Format labels
  labs(title = "Covid-19 Variants in California", x = "",
       y = "Percentage of sequenced specimens", color = "") +
  #Format x-axis to have abbreviated month and full year labels and increment
  #by month
  scale_x_date(date_labels = "%b %Y",
               breaks = function(x) seq.Date(from = as.Date("2021-01-01"),
                                             to = max(x), by = "1 month")) +
  #Adjust position and angle of x-axis labels
  theme(axis.text.x = element_text(angle = 60, hjust = 1))
```

Covid−19 Variants in California